Bivariate Splines for Spatial Functional Regression Models

Serge Guillas
Department of Statistical Science, University College London,
London, WC1E 6BTS, UK.
serge@stats.ucl.ac.uk

Ming-Jun Lai
Department of Mathematics, The University of Georgia
Athens, GA 30602, USA.
mjlai@math.uga.edu

Abstract

We consider the functional linear regression model where the explanatory variable is a random surface and the response is a real random variable, with bounded or normal noise. Bivariate splines over triangulations represent the random surfaces. We use this representation to construct least squares estimators of the regression function with or without a penalization term. Under the assumptions that the regressors in the sample are bounded and span a large enough space of functions, bivariate splines approximation properties yield the consistency of the estimators. Simulations demonstrate the quality of the asymptotic properties on a realistic domain. We also carry out an application to ozone concentration forecasting over the US that illustrates the predictive skills of the method.

1 Introduction

In various fields, such as environmental science, finance, geological science and biological science, large data sets are becoming readily available, e.g., by real time monitoring such as satellites circulating around the earth. Thus, the objects of statistical study are curves, surfaces and manifolds, in addition to the traditional points, numbers or vectors. Functional Data Analysis (FDA) can help represent and analyze infinite-dimensional random processes [17, 10]. FDA aggregates consecutive discrete recordings and views them as sampled values of a random curve or random surface, keeping track of order or smoothness. In this context, random curves have been the focus on many studies, but very few address the case of surfaces.

In regression, when the explanatory variable is a random function and the response is a real random variable, we can define a so-called functional linear model, see Chapter 15 in [17] and references therein. In particular, [4] and [5] introduced consistent estimates based on functional principal components, and decompositions in univariate splines spaces. The model can be generalized to the bivariate setting as follows. Let Y be a real-valued random variable. Let $\mathcal{D}$ be a polygonal domain in $\mathbf{R}^2$. The regression model is:

$$Y = f(X) = \langle g, X \rangle = \int_{\mathcal{D}} g(s)X(s)ds + \varepsilon, \tag{1}$$

where $g(s)$ is in a function space H (usually $= L^2(\mathcal{D})$), ε is a real random variable that satisfies $E\varepsilon = 0$ and $EX(s)\varepsilon = 0, \forall s \in \mathcal{D}$. One of the objectives in FDA is to determine or approximate g which is defined on a 2D spatial domain $\mathcal{D}$ from the observations on X obtained over a set of design points in $\mathcal{D}$ and Y.

This model in the univariate setting has been extensively studied using many different approaches. When the curves are supposed to be fully observed, it is possible to use the Karhunen-Loève expansion, or principal components analysis for curves [3, 20, 12]. However, as pointed out by [13], when the curves are not fully observed, which is obviously the case in practice, FDA would then proceed as though some smooth approximation of the true curves were the true curves. One typical approach is based on univariate splines [5, 6, 7], whereas [12] and [3] use a local-linear smoother, which helps derive asymptotic results. [5] introduced the Penalized B-splines estimator (PS) and the Smooth Principal Component Regression (SPCR) estimator in one dimension. For the consistency of these estimates, they either assume that the noise is bounded (PS) or use hypotheses on the rate of decay of eigenvalues for the covariance operator of X (SPCR).

Motivated by the studies mentioned above, we study here the similar problem in the two-dimensional setting, i.e., a functional regression model where the explanatory variable is a random surface and the response is a real random variable. To express a random surface over 2D irregular polygonal domain $\mathcal{D}$, we shall use bivariate splines which are smooth piecewise polynomial functions over a 2D triangulated polygonal domain $\mathcal{D}$. They are similar to univariate splines defined on piecewise subintervals. The theory of such bivariate spline functions are recently matured, see a monograph [16]. For example, we know the approximation properties of bivariate spline spaces and construction of locally supported bases. Computational algorithms for scattered data interpolation and fitting are available in [1]. In particular, computing integrals with bivariate splines is easy, so it is now possible to use bivariate splines to build regression models for random surfaces. Certainly, it is possible to use the standard finite element method or thin-plate spline method for functional data analysis as in [18] and in [19] in a non-functional context. A finite element (FE) analysis was carried out for smoothing the data over complicated domains in [18] and thin-plate splines are used in regression in [19]. In addition, it is also possible to use a tensor product of univariate splines or wavelets when the domain of interest is rectangular. We find that our spline method is particularly easy to use, and hence will be used in our numerical experiments to be reported in the last section. We shall leave the investigation of using finite element method, thin-plate spline method, and tensor product of univariate B-splines or wavelets for 2D FDA to the interested reader.

Our approach to FDA in the bivariate setting is a straightforward approach which is different from the approaches in [4, 5, 6, 7]. Mainly we use the fact that bivariate spline space can be dense in the standard $L_2(\mathcal{D})$ space and many other spaces. We can approximate g and X in (1) using spline functions and build a regression model. In our approach, we may assume that the noise is Gaussian or bounded and we do not make explicit assumptions on the covariance structure of X. The only requirement in our approach is that all the random functions X span a large enough space so that g can be well estimated. It is clear that this is a reasonable assumption. In this paper, we mainly derive rates for convergence in probability for S_g, a spline approximation of g and the two empirical estimates when using bivariate splines to approximate X using discrete least squares method and penalized least squares method. This convergence of empirical estimates of S_g to g in L_2 norm is currently under investigation by the authors. We have implemented our approach using bivariate splines and perform numerical simulation and forecasting experiments with a set of real data. Comparison with univariate forecasting methods are given to show that our approach works very well. To our knowledge, our paper is the first piece of work on functional regression of a real random variable onto random surfaces.

The paper is organized as follows. After introducing bivariate splines in the preliminary section, we consider approximations of linear functionals with or without penalty term in the next two sections. Then we address the case of discrete observations of random surfaces in section 5. In order to illustrate the findings on an irregular region, in section 6 we carry out simulations, and forecasting with real data, for which the domain is delimited by the United States frontiers, and the sample points are the US EPA monitoring locations. Our numerical experiments demonstrate the efficiency and convenience of using bivariate splines to approximate linear functionals in functional data regression analysis.

2 Preliminary on Bivariate Splines

Let $\mathcal{D}$ be a polygonal domain in $\mathbf{R}^2$. Let $\triangle$ be a triangulation of $\mathcal{D}$ in the following sense: $\triangle$ is a collection of triangles $t \subset \mathcal{D}$ such that $\cup_{t\in\triangle} t = \mathcal{D}$ and the intersection of any two triangles $t_1, t_2 \in \triangle$ is either an empty set or their common edge of t_1, t_2 or their common vertex of t_1, t_2. For each $t \in \triangle$, let $|t|$ denote the longest length of the edges of t, and $|\triangle|$ the size of triangulation, which is the longest length of the edges of $\triangle$. Let $\theta_\triangle$ denote the smallest angle of $\triangle$. Next let $S_d^r(\triangle) = \{h \in C^r(\mathcal{D}), h|_t \in \mathbf{P}_d, t \in \triangle\}$ be the space of all piecewise polynomial functions h of degree d and smoothness r over $\triangle$, where $\mathbf{P}_d$ is the space of all polynomials of degree d. Such spline spaces have been studied in depth in the last twenty years and a basic theory and many important results are summarized in [16]. Throughout the paper, $d \geq 3r + 2$. Then it is known [15, 16] that the spline space $S_d^r(\triangle)$ possesses an optimal approximation property: Let D_1 and D_2 denote the derivatives with respect to the first and second variables, $\|h\|_{L_p(\mathcal{D})}$ stand for the usual L_p norm of f over $\mathcal{D}$, $|h|_{m,p,\mathcal{D}}$ the L_p norm of the m^{th} derivatives of h over $\mathcal{D}$, and $W_p^{m+1}(\mathcal{D})$ be the usual Sobolev space over $\mathcal{D}$.

Theorem 2.1 *Suppose that $d \geq 3r+2$ and $\triangle$ be a triangulation. Then there exists a quasi-interpolatory operator $Qh \in S_d^r(\triangle)$ mapping any $h \in L_1(\mathcal{D})$ into $S_d^r(\triangle)$ such that Qh achieves the optimal approximation order: if $h \in$*

$W_p^{m+1}(\mathcal{D})$,

$$\|D_1^\alpha D_2^\beta(Qh-h)\|_{L_p(\mathcal{D})} \le C|\triangle|^{m+1-\alpha-\beta}|h|_{m+1,p,\mathcal{D}} \tag{2}$$

for all $\alpha+\beta \le m+1$ with $0 \le m \le d$, where C is a constant which depends only on d and the smallest angle $\theta_\triangle$ and may be dependent on the Lipschitz condition of the boundary of $\mathcal{D}$.

Bivariate splines have been used for scattered data fitting and interpolation for many years. Typically, the minimal energy spline interpolation, discrete least squares splines for data fitting and penalized least squares splines for data smoothing are used. Their approximation properties have been studied and numerical algorithms for these data fitting methods have been implemented and tested. See [1] and [14] and the references therein.

3 Approximation of Linear Functionals

In this section we use a spline space $S_d^r(\triangle)$ with smoothness $r > 0$ and degree $d \ge 3r+2$ over a triangulation $\triangle$ of a bounded domain $\mathcal{D} \subset \mathbf{R}^2$ with $|\triangle| < 1$ sufficiently small. The triangulation is fixed and thus the spline basis and its cardinality m as well. We propose a new approach to study approximation of the given functional f on the random functions X taking their values in H. Here H is a Hilbert space, for example, $H = W_2^\nu(\mathcal{D})$, the standard Sobolev space of all ν^{th} differentiable functions which are square integrable over $\mathcal{D}$ for an integer $\nu \ge r > 0$, where r is the smoothness of our spline space $S_d^r(\triangle)$.

We assume that X and Y follow the regression model (1). Then α that solves the following minimization problem is equal to g:

$$\alpha = \min_{\beta \in H} E\left[(f(X) + \epsilon - \langle \beta, X\rangle)^2\right]. \tag{3}$$

Since $S_d^r(\triangle)$ can be dense in H as $|\triangle| \to 0$ based on Theorem 2.1, we look for an approximation $S_\alpha \in S_d^r(\triangle)$ of α such that

$$S_\alpha = \arg\min_{\beta \in S_d^r(\triangle)} E\left[(f(X) + \epsilon - \langle \beta, X\rangle)^2\right]. \tag{4}$$

We now begin to analyze how S_α approximates α in terms of the size $|\triangle|$ of triangulation.

Let $\{\phi_1, \cdots, \phi_m\}$ be a basis for $S_d^r(\triangle)$. We write $S_\alpha = \sum_{j=1}^m c_j \phi_j$. Then, by directly calculating the least squares solution, the coefficient vector $\mathbf{c} = (c_1, \cdots, c_m)^T$ of S_α satisfies the following linear system

$$A\mathbf{c} = \mathbf{b}$$

with A being a matrix of size $m \times m$ whose entries are $E(\langle \phi_i, X \rangle \langle \phi_j, X \rangle)$ for $i, j = 1, \cdots, m$ and $\mathbf{b}$ being a vector of length m with entries $E((f(X) + \epsilon)\langle \phi_j, X \rangle) = E(f(X)\langle \phi_j, X \rangle)$ for $j = 1, \cdots, m$. Note that A is the matrix representation of the covariance function of X exercised at the elements $\phi_j, j = 1, \cdots, m$. Although we do not know how $X \in H$ is distributed, let us assume that only the zero spline function in $S_d^r(\triangle)$ is orthogonal to all functions in the collection $\mathcal{X} = X(\omega), \omega \in \Omega$, where Ω is the probability space. In this case, we claim that A is invertible. Otherwise, we would have $\mathbf{c}^T A \mathbf{c} = 0$, i.e., $E(\langle \sum_{i=1}^m c_i \phi_i, X \rangle)^2 = 0$. That is, $\sum_{i=1}^m c_i \phi_i$ is orthogonal to X for all $X(\omega) \in \mathcal{X}$. It follows from the assumption that $\sum_{i=1}^m c_i \phi_i$ is a zero spline and hence, $c_i = 0$ for all i. Thus, we have obtained the following

Theorem 3.1 *Suppose that only the zero spline in $S_d^r(\triangle)$ is orthogonal to the collection $\mathcal{X} \subset H$. Then the minimization problem (4) has a unique solution in $S_d^r(\triangle)$.*

To see that S_α is a good approximation of α, we let $\{\phi_j, j = m + 1, m + 2, \cdots, \}$ be a basis of the orthogonal complement space of $S_d^r(\triangle)$ in H. Then we can write $\alpha = \sum_{j=1}^\infty c_j \phi_j$. Note that the minimization in (3) yields $E(\langle \alpha, X \rangle \langle \phi_j, X \rangle) = E(f(X)\langle \phi_j, X \rangle)$ for all $j = 1, 2, \cdots$ while the minimization in (4) gives

$$E(\langle S_\alpha, X \rangle \langle \phi_j, X \rangle) = E(f(X)\langle \phi_j, X \rangle)$$

for all $j = 1, 2, \cdots, m$. It follows that

$$E(\langle \alpha - S_\alpha, X \rangle \langle \phi_j, X \rangle) = 0 \tag{5}$$

for all $j = 1, 2, \cdots, m$. Let Q_α be the quasi-interpolatory spline in $S_d^r(\triangle)$ which achieves the optimal order of approximation of α from $S_d^r(\triangle)$ as in Preliminary section. Then (5) implies that

$$\begin{aligned} E((\langle \alpha - S_\alpha, X \rangle)^2) &= E(\langle \alpha - S_\alpha, X \rangle \langle \alpha - Q_\alpha, X \rangle) \\ &\le (E((\langle \alpha - S_\alpha, X \rangle)^2))^{1/2} E((\langle \alpha - Q_\alpha, X \rangle)^2)^{1/2}. \end{aligned}$$

It follows that $E((\langle \alpha - S_\alpha, X \rangle)^2) \le E((\langle \alpha - Q_\alpha, X \rangle)^2) \le \|\alpha - Q_\alpha\|_H^2 E(\|X\|^2)$. The approximation of the quasi-interpolant Q_α of α (cf. Theorem 2.1) gives:

Theorem 3.2 *Suppose that* $E(\|X\|^2) < \infty$ *and suppose* $\alpha \in C^\nu(\mathcal{D})$ *for* $r \le \nu \le d+1$. *Then the solution* S_α *from the minimization problem (4) approximates* α *in the sense:* $E((\langle \alpha - S_\alpha, X\rangle)^2) \le C|\triangle|^{2\nu} E(\|X\|^2)$, *where* $|\triangle|$ *is the maximal length of the edges of* $\triangle$.

Next we consider the empirical estimate of S_α. Let $X_i, i = 1, \cdots, n$ be a sequence of i.i.d functional random variables such that only the zero spline function in the space $S^r_d(\triangle)$ is perpendicular to the subspace spanned by $\{X_1, \cdots, X_n\}$ except on an event whose probability p_n goes to zero as $n \to +\infty$. The empirical estimate $\widehat{S_{\alpha,n}} \in S^r_d(\triangle)$ is the solution of

$$\widehat{S_{\alpha,n}} = \arg \min_{\beta \in S^r_d(\triangle)} \frac{1}{n} \sum_{i=1}^{n} (f(X_i) + \epsilon_i - \langle \beta, X_i\rangle)^2. \tag{6}$$

In fact the solution of the above minimization is given by $\widehat{S_{\alpha,n}} = \sum_{i=1}^m c_{n,i}\phi_i$ with coefficient vector $\mathbf{c}_n = (c_{n,i}, i = 1, \cdots, m)$ satisfying $\widehat{A}_n \mathbf{c}_n = \widehat{b}_n$,
$\widehat{A}_n = \left[\frac{1}{n}\sum_{\ell=1}^n \langle \phi_i, X_\ell\rangle\langle \phi_j, X_\ell\rangle\right]_{i,j=1,\cdots,m}$ and
$\widehat{b}_n = \left[\frac{1}{n}\sum_{\ell=1}^n f(X_\ell)\langle \phi_j, X_\ell\rangle + \frac{1}{n}\sum_{\ell=1}^n \langle \phi_j, \epsilon_\ell X_\ell\rangle\right]_{j=1,\cdots,m}$.

We have the following:

Theorem 3.3 *Suppose that only the zero spline function in the spline space* $S^r_d(\triangle)$ *is perpendicular to the subspace span*$\{X_1, \cdots, X_n\}$ *except on an event whose probability* p_n *goes to zero as* $n \to +\infty$. *Then, with probability* $1 - p_n$, *there exists a unique* $\widehat{S_{\alpha,n}} \in S^r_d(\triangle)$ *minimizing (6).*

Proof. It is straightforward to see that the coefficient vector of $\widehat{S_{\alpha,n}}$ satisfies the above relations. To see that $\widehat{A}_n \mathbf{c}_n = \widehat{b}_n$ has a unique solution, we claim that if $\widehat{A}_n c' = 0$, then $c' = 0$. It follows that $(c')^T \widehat{A}_n c' = 0$, i.e., $\sum_{\ell=1}^n (\langle \sum_{i=1}^m c'_i \phi_i, X_\ell\rangle)^2 = 0$. That is, $\sum_{i=1}^m c'_i\phi_i$ is orthogonal to $X_\ell, \ell = 1, \cdots, n$. According to the assumption, $c' = 0$ except for an event whose probability p_n goes to zero when $n \to +\infty$. ■

We now prove that $\widehat{S_{\alpha,n}}$ approximates S_α in probability. To this end we need the following lemmas.

Lemma 3.1 *Suppose that* $\triangle$ *is a* β*-quasi-uniform triangulation (cf. [16]). There exist two positive constants* C_1 *and* C_2 *independent of* $\triangle$ *such that for any spline function* $S \in S^r_d(\triangle)$ *with coefficient vector* $\mathbf{s} = (s_1, \cdots, s_m)^T$ *with* $S = \sum_{i=1}^m s_j\phi_j$, $C_1|\triangle|^2\|\mathbf{s}\|^2 \le \|S\|^2 \le C_2|\triangle|^2\|\mathbf{s}\|^2$.

A proof of this lemma can be found in [15] and [16]. The following lemma is well-known in numerical analysis [[11],p.82].

Lemma 3.2 *Let A be an invertible matrix and $\tilde{A}$ be a perturbation of A satisfying $\|A^{-1}\|\,\|A-\tilde{A}\| < 1$. Suppose that x and $\tilde{x}$ are the exact solutions of $Ax = b$ and $\tilde{A}\tilde{x} = \tilde{b}$, respectively. Then*

$$\frac{\|x-\tilde{x}\|}{\|x\|} \leq \frac{\kappa(A)}{1-\kappa(A)\frac{\|A-\tilde{A}\|}{\|A\|}} \left[\frac{\|A-\tilde{A}\|}{\|A\|} + \frac{\|b-\tilde{b}\|}{\|b\|}\right].$$

Here, $\kappa(A) = \|A\|\|A^{-1}\|$ denotes the condition number of matrix A.

We need to analyze the differences between the entries of A and $\widehat{A}_n$ as well as the differences between $\mathbf{b}$ and $\widehat{b}_n$. The standard assumptions of the strong Law of Large Numbers are satisfied for the sequence $(\langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle)$, $\ell = 1, 2, ..., n$. Indeed, the sequence $\{\langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle, \ell = 1, 2, ..., n\}$ is i.i.d, the sequence $(\langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle)$, $\ell = 1, 2, ...$ is integrable and $E(\langle\phi_i, X_l\rangle\langle\phi_j, X_l\rangle) \leq E(\|\phi_i\|\|\phi_j\|\|X_l\|^2)$ by Cauchy-Schwarz's inequality. Let us assume $E(\|X_l\|^2) \leq B < \infty$. Then we have $E(\|\phi_i\|\|\phi_j\|\|X_l\|^2) \leq B^2\|\phi_i\|\|\phi_j\| < \infty$, because the basis spline functions ϕ_j can be chosen to be bounded in $L_2(\mathcal{D})$ for all j independent of triangulation $\triangle$ [16]. Thus by the Strong Law of Large Number, for each i, j, we have $\frac{1}{n}\sum_{\ell=1}^n \langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle - E(\langle\phi_i, X\rangle\langle\phi_j, X\rangle) \to 0$ almost surely.

Not only we know that $\widehat{A}_n$ converge to A entrywisely, but also we can give a global rate of convergence. Let $\xi_l = \langle\phi_i, X_l\rangle\langle\phi_j, X_l\rangle$. The ξ_l are bounded. Let $M = \max_{ij}\max_\ell |\langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle| \leq \max_{ij}\max_\ell \|\phi_i\|\|\phi_j\|\|X_\ell\|^2$. For each i, $|\xi_l| \leq M < \infty$ almost surely. So we can find a rate of convergence by applying the following Hoeffding's exponential inequality [2, p. 24]

Lemma 3.3 *Let $\{\xi_l\}_{l=1}^n$ be n independent random variables. Suppose that there exists a positive number M such that for each i, $|\xi_l| \leq M < \infty$ almost surely. Then $P\left(|\frac{1}{n}\sum_{\ell=1}^n(\xi_l - E(\xi_l))| \geq \delta\right) \leq 2\exp\left(-\frac{n\delta^2}{2M^2}\right)$ for $\delta > 0$.*

To use Lemma 3.2, we use the maximum norm for matrix $A - \widehat{A}_n$ and vector $b - \widehat{b}_n$. For simplicity, let us write $[a_{ij}]_{1\leq i,j\leq m} = A - \widehat{A}_n$ and hence,

$$P\left(\|[a_{ij}]_{1\leq i,j\leq m}\|_\infty \geq \delta\right)$$

$$
\begin{aligned}
&= P\left(\max_{1\le i\le m}\sum_{j=1}^m |a_{ij}|\ge\delta\right)\le\sum_{i=1}^m P\left(\sum_{j=1}^m|a_{ij}|\ge\delta\right)\\
&\le \sum_{i=1}^m\sum_{j=1}^m P\left(|a_{ij}|\ge\delta/m)\right)\le 2m^2\exp\left(-\frac{n\delta^2}{2m^2M^2}\right), \qquad (7)
\end{aligned}
$$

using Lemma 3.3.

Similarly, we can estimate the entries of $\mathbf{b}-\widehat{b}_n$. We denote its entries by $b_j=-\frac{1}{n}\sum_{\ell=1}^n f(X_\ell)\langle\phi_j,X_\ell\rangle-E(f(X)\langle\phi_j,X\rangle)+\frac{1}{n}\sum_{\ell=1}^n\langle\phi_j,\epsilon_\ell X_\ell\rangle$. Let us write $b_j=b_j^1+b_j^2$ with b_j^1 and b_j^2 being the first and second terms, respectively. It is easy to see that $P(|b_j|\ge\delta)\le P(|b_j^1|\ge\delta/2)+P(|b_j^2|\ge\delta/2)$. Since the functional f is bounded, $|f(X_\ell)|\le F\|X_\ell\|$, with a finite constant F. By Lemma 3.3, we have $P\left(|b_j^1|\ge\delta/2\right)\le 2\exp\left(-\frac{n\delta^2}{8M_b^2}\right)$, where $M_b=\max_j|f(X_\ell)\langle\phi_j,X_\ell\rangle|\le F\|X_\ell\|\|\phi_j\|\,\|X_\ell\|$ which is a finite quantity since $\|X_\ell\|$ is bounded almost surely. Regarding the second term b_j^2, we first consider the case where the random noises ϵ_ℓ are bounded almost surely. We will consider the case where ϵ_ℓ are Gaussian noises just after. Letting $\xi_\ell=\langle\phi_j,\epsilon_\ell X_\ell\rangle$, we apply Lemma 3.3 to have $P\left(|b_j^2|\ge\delta/2\right)\le 2\exp\left(-\frac{n\delta^2}{8M_\epsilon^2}\right)$ where $M_\epsilon=\max_j|\langle\phi_j,\epsilon_\ell X_\ell\rangle|\le\max_j\|\phi_j\||\epsilon_\ell|\|X_\ell\|$ which is finite under the assumption that both $\|X_\ell\|$ and $|\epsilon_\ell|$ are bounded almost surely.

Thus we have

$$
P\left(\|\mathbf{b}-\widehat{b}_n\|_\infty\ge\delta\right)\le\sum_{j=1}^m P\left(|b_j|\ge\delta\right)\le 2m\exp\left(-\frac{n\delta^2}{8M_b^2}\right)+2m\exp\left(-\frac{n\delta^2}{8M_\epsilon^2}\right). \qquad (8)
$$

Therefore we combine the estimates (7) and (8) and use Lemmas 3.1 and 3.2 to give an estimate of the convergence of $\widehat{S_{\alpha,n}}$ to S_α in probability. We first use Lemma 3.1 to get $P\left(\frac{\|S_\alpha-\widehat{S_{\alpha,n}}\|}{\|S_\alpha\|}\ge\delta\right)\le P\left(\frac{\|\mathbf{c}-\widehat{c}_n\|}{\|\mathbf{c}\|}\ge\gamma\delta\right)$ where $\gamma=\sqrt{\dfrac{C_1}{C_2}}$. Next we use Lemma 3.2. For simplicity we let $\beta=\frac{\|\mathbf{c}-\widehat{c}_n\|}{\|\mathbf{c}\|},\eta=\frac{\|A-\widehat{A}_n\|}{\|A\|}$ and $\theta=\frac{\|\mathbf{b}-\widehat{b}_n\|}{\|\mathbf{b}\|}$. Then Lemma 3.2 implies that

$$
\begin{aligned}
& P\left(\beta\ge\gamma\delta\right)\\
\le\;& P\left(\beta\ge\gamma\delta,\kappa(A)\eta\le 1/2\right)+P\left(\beta\ge\gamma\delta,\kappa(A)\eta\ge 1/2\right)\\
\le\;& P\left(\frac{\kappa(A)}{1-\kappa(A)\eta}(\eta+\theta)\ge\gamma\delta,\kappa(A)\eta\le 1/2\right)+P\left(\kappa(A)\eta\ge 1/2\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq P\left((\eta+\theta) \geq \frac{\gamma\delta}{2\kappa(A)}\right) + P\left(\kappa(A)\eta \geq 1/2\right) \\
&\leq P\left(\eta \geq \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\theta \geq \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\eta \geq \frac{\gamma\delta}{2\kappa(A)}\right) \\
&\leq 2P\left(\eta \geq \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\theta \geq \frac{\gamma\delta}{4\kappa(A)}\right)
\end{aligned}
$$

for all $\delta \leq 1$. We combine the estimates (7) and (8) to get

$$
\begin{aligned}
P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,n}}\|}{\|S_\alpha\|} \geq \delta\right) &\leq 4m^2 \exp\left(-\frac{n\gamma^2\delta^2}{32\kappa(A)^2 m^2 M^2}\right) \\
&+2m \exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2 M_b^2}\right) + 2m \exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2 M_\epsilon^2}\right). \qquad (9)
\end{aligned}
$$

Therefore we have obtained one of the major theorems in this section.

Theorem 3.4 *Suppose that X_ℓ, $\ell = 1, \cdots, n$ are independent and identically distributed and X_1 is bounded almost surely. Suppose that the ϵ_ℓ are independent and bounded almost surely. Assume that $f(X)$ is a bounded linear functional. Then $\widehat{S_{\alpha,n}}$ converges to S_α in probability with convergence rate in (9).*

As an example, if we choose $m = n^{1/4}$, we get a convergence rate of $n^{1/2} \exp\left(-\frac{\sqrt{n}\gamma^2\delta^2}{32\kappa(A)^2 M^2}\right)$ which is the slower of the terms.

We are now ready to consider the case where ϵ_ℓ is a Gaussian noise $N(0, \sigma_\ell^2)$ for $\ell = 1, \cdots, n$. Instead of Lemma 3.3, it is easy to prove

Lemma 3.4 *Suppose that ϵ_ℓ is a Gaussian noise $N(0, \sigma_\ell^2)$ for $\ell = 1, \cdots, n$. Then*

$$
P(|\frac{1}{n}\sum_{\ell=1}^{n} \epsilon_\ell| > \delta) \leq \exp(-\frac{n^2\delta^2}{2\sum_{\ell=1}^{n}\sigma_\ell^2}).
$$

It follows that when ϵ_ℓ are independent and identically distributed Gaussian noises $N(0, \sigma^2)$, $P\left(|\frac{1}{n}\sum_{\ell=1}^{n}(\epsilon_\ell Y_\ell))| \geq \delta\right) \leq \exp\left(-\frac{n\delta^2}{2\sigma^2 C^2}\right)$for $\delta > 0$, under the assumption that Y_ℓ are independent random variables which are bounded by C, i.e., $\|Y_\ell\| \leq C$. Similar to the proof of Theorem 3.4, with $Y_\ell = \langle \phi_j, X_\ell \rangle$ in that case, we have

$$
P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,n}}\|}{\|S_\alpha\|} \geq \delta\right) \leq 4m^2 \exp\left(-\frac{n\gamma^2\delta^2}{32\kappa(A)^2 m^2 M^2}\right)
$$

$$+2m\exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2M_b^2}\right)+2m\exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2\sigma^2C^2}\right). \quad (10)$$

Theorem 3.5 *Suppose that X_ℓ, $\ell = 1, \cdots, n$ are independent and identically distributed random variables and X_1 is bounded almost surely. Suppose ϵ_ℓ are independent and identically distributed Gaussian noise $N(0, \sigma^2)$ and $f(X)$ is a bounded linear functional. Then $\widehat{S_{\alpha,n}}$ converges to S_α in probability with convergence rate in (10).*

4 Approximation of Linear Functionals with Penalty

In this section we propose another new approach to study the functional f in model (1). Mainly we seek a solution $\alpha \in H$ which solves the following minimization problem:

$$\alpha = \arg\min_{\beta\in H} E\left[(f(X) + \epsilon - \langle\beta, X\rangle)^2\right] + \rho\|\beta\|_r^2, \quad (11)$$

where $\rho > 0$ is a parameter and $\|\beta\|_r^2$ denotes the semi-norm of β: $\|\beta\|_r^2 = \mathcal{E}_r(\beta,\beta)$, where $\mathcal{E}_r(\alpha,\beta) = \int_{\mathcal{D}} \sum_{k=0}^r \sum_{i+j=k} D_1^i D_2^j \alpha D_1^i D_2^j \beta$, and D_1 and D_2 stand for the partial derivatives with respect to the first and second variables. Contrarily to the previous section, α is not necessarily equal to g. Since $S_d^r(\triangle)$ can be dense in H as $|\triangle| \to 0$, we consider a spline space $S_d^r(\triangle)$ for a smoothness $r \geq 0$ and degree $d > r$ over a triangulation $\triangle$ of $\mathcal{D}$ with $|\triangle|$ sufficiently small. Note that in this section, as in the previous one, the triangulation is fixed and thus the spline basis and its cardinality m as well. We look for an approximation $S_{\alpha,\rho} \in S_d^r(\triangle)$ of α such that

$$S_{\alpha,\rho} = \arg\min_{\beta\in S_d^r(\triangle)} E\left[(f(X) + \epsilon - \langle\beta, X\rangle)^2\right] + \rho\mathcal{E}_r(\beta). \quad (12)$$

We now analyze how $S_{\alpha,\rho}$ approximates α in terms of the size $|\triangle|$ of triangulation and $\rho \to 0+$. Again let $\{\phi_1, \cdots, \phi_m\}$ be a basis for $S_d^r(\triangle)$. We write $S_\alpha = \sum_{j=1}^m c_j\phi_j$. Then a direct calculation of the least squares solution of (12) entails that the coefficient vector $\mathbf{c} = (c_1, \cdots, c_m)^T$ satisfies a linear system $A\mathbf{c} = \mathbf{b}$ with A being a matrix of size $m \times m$ with entries $E(\langle\phi_i, X\rangle\langle\phi_j, X\rangle) + \rho\mathcal{E}_r(\phi_i, \phi_j)$ for $i, j = 1, \cdots, m$ and $\mathbf{b}$ being a vector of length m with entries $E(f(X)\langle\phi_j, X\rangle)$ for $j = 1, \cdots, m$.

Although we do not know how $X \in H$ is distributed, let us assume that only the zero polynomial is orthogonal to all functions in the collection $\mathcal{X} = X(\omega), \omega \in \Omega$ in the standard Hilbert space $L_2(\mathcal{D})$. In this case, A is invertible. Otherwise, we would have $\mathbf{c}^T A\mathbf{c} = 0$, i.e.,

$$E(\langle \sum_{i=1}^{m} c_i\phi_i, X\rangle)^2 + \rho \| \sum_{i=1}^{m} c_i\phi_i\|_r^2 = 0 \tag{13}$$

Since the second term in (13) is equal to zero, $\sum_{i=1}^{m} c_i\phi_i$ is a polynomial of degree $< r$. As the first term in (13) is also zero, this polynomial is orthogonal to X for all $X \in \mathcal{X}$. By the assumption, $\sum_{i=1}^{m} c_i\phi_i$ is a zero spline and hence, $c_i = 0$ for all i. Thus, we have obtained the following

Theorem 4.1 *Suppose that only the zero polynomial is orthogonal to the collection $\mathcal{X}$ in $L_2(\mathcal{D})$. Then the minimization problem (12) has a unique solution in $S_d^r(\triangle)$.*

To see that $S_{\alpha,\rho}$ is a good approximation of α, we let $\{\phi_j, j = m+1, m+2, \cdots\}$ be a basis of the orthogonal complement space of $S_d^r(\triangle)$ in $L_2(\mathcal{D})$. Then we can write $\alpha = \sum_{j=1}^{\infty} c_j\phi_j$. Note that the minimization in (11) yields $E(\langle\alpha, X\rangle\langle\phi_j, X\rangle) + \rho\mathcal{E}_r(\alpha, \phi_j) = E(f(X)\langle\phi_j, X\rangle)$ for all $j = 1, 2, \cdots$ while the minimization in (12) gives

$$E(\langle S_\alpha, X\rangle\langle\phi_j, X\rangle) + \rho\mathcal{E}_r(S_\alpha, \phi_j) = E(f(X)\langle\phi_j, X\rangle)$$

for all $j = 1, 2, \cdots, m$. It follows that

$$E(\langle\alpha - S_{\alpha,\rho}, X\rangle\langle\phi_j, X\rangle) + \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, \phi_j) = 0 \tag{14}$$

for $j = 1, \cdots, m$. Let Q_α be the quasi-interpolatory spline in $S_d^r(\triangle)$ which achieves the optimal order of approximation of α from $S_d^r(\triangle)$ as in Preliminary section. Then (14) implies that

$$\begin{aligned} E((\langle\alpha - S_{\alpha,\rho}, X\rangle)^2) &= E(\langle\alpha - S_{\alpha,\rho}, X\rangle\langle\alpha - Q_\alpha, X\rangle) - \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, Q_\alpha - S_{\alpha,\rho}) \\ &\le (E((\langle\alpha - S_{\alpha,\rho}, X\rangle)^2))^{1/2} E((\langle\alpha - Q_\alpha, X\rangle)^2)^{1/2} \\ &\quad -\rho\|\alpha - S_{\alpha,\rho}\|_r^2 + \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, \alpha - Q_\alpha) \\ &\le \frac{1}{2}E((\langle\alpha - S_{\alpha,\rho}, X\rangle)^2) + \frac{1}{2}E((\langle\alpha - Q_\alpha, X\rangle)^2) \\ &\quad -\frac{1}{2}\rho\|\alpha - S_{\alpha,\rho}\|_r^2 + \frac{1}{2}\rho\|\alpha - Q_\alpha\|_r^2. \end{aligned}$$

Hence $E((\langle\alpha - S_{\alpha,\rho}, X\rangle)^2) + \rho\|\alpha - S_{\alpha,\rho}\|_r^2 \le E((\langle\alpha - Q_\alpha, X\rangle)^2) + \rho\|\alpha - Q_\alpha\|_r^2$. The approximation of the quasi-interpolant Q_α of α [15] gives:

Theorem 4.2 *Suppose that* $E(\|X\|^2) < \infty$ *and suppose* $\alpha \in C^\nu(\mathcal{D})$ *for* $\nu \geq r$. *Then the solution* $S_{\alpha,\rho}$ *from the minimization problem (12) approximates* α: $E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2) \leq C|\triangle|^{2\nu} E(\|X\|^2) + \rho C|\triangle|^{2(\nu-r)}$ *where* C *is a positive constant independent of the size* $|\triangle|$ *of triangulation* $\triangle$.

Next we consider the empirical estimate of $S_{\alpha,\rho}$. Let $X_i, i = 1, \cdots, n$ be a sequence of functional random variables such that only the zero polynomial is perpendicular to the subspace spanned by $\{X_1, \cdots, X_n\}$ except on an event whose probability p_n goes to zero as $n \to +\infty$. The empirical estimate $\widehat{S_{\alpha,\rho,n}} \in S_d^r(\triangle)$ is the solution of

$$\widehat{S_{\alpha,\rho,n}} = \arg \min_{\beta \in S_d^r(\triangle)} \frac{1}{n} \sum_{i=1}^n (f(X_i) + \epsilon_i - \langle \beta, X_i\rangle)^2 + \rho \|\beta\|_r^2, \qquad (15)$$

with $\rho > 0$ the smoothing parameter. The solution of the above minimization is given by $\widehat{S_{\alpha,\rho,n}} = \sum_{i=1}^m c_{n,i}\phi_i$ with coefficient vector $\mathbf{c}_n = (c_{n,i}, i = 1, \cdots, m)$ satisfying $\widehat{A}_n \mathbf{c}_n = \widehat{b}_n$, where

$$\widehat{A}_n = \left[\frac{1}{n}\sum_{\ell=1}^n \langle \phi_i, X_\ell\rangle \langle \phi_j, X_\ell\rangle + \rho \mathcal{E}_r(\phi_i, \phi_j)\right]_{i,j=1,\cdots,m}$$

and

$$\widehat{b}_n = \left[\frac{1}{n}\sum_{\ell=1}^n (f(X_\ell) + \epsilon_\ell)\langle \phi_j, X_\ell\rangle\right]_{j=1,\cdots,m}.$$

Similar to the proof of the Theorems in the previous section, we can prove the following

Theorem 4.3 *Suppose that only the zero polynomial is perpendicular to the subspace spanned by* $\{X_1, \cdots, X_n\}$ *except on an event whose probability* p_n *goes to zero as* $n \to +\infty$. *Then there exists a unique* $\widehat{S_{\alpha,\rho,n}} \in S_d^r(\triangle)$ *minimizing (15) with probability* $1 - p_n$.

We now prove that $\widehat{S_{\alpha,\rho,n}}$ approximates $S_{\alpha,\rho}$ in probability. The entries of $A - \widehat{A}_n$ and $\mathbf{b} - \widehat{b}$ are exactly the same as those in the previous section. Hence, the same analysis as in the previous section yields

Theorem 4.4 *Suppose that* X_ℓ, $\ell = 1, \cdots, n$ *are independent and identically distributed random variables and* X_1 *is bounded almost surely. Suppose that* ϵ_ℓ

are independent and identically distributed random noises which are bounded almost surely. Suppose that $f(X_\ell)$ is a bounded linear functional. Then $\widehat{S_{\alpha,\rho,n}}$ converges to $S_{\alpha,\rho}$ in probability with convergence rate in (9).

Theorem 4.5 *Suppose that X_ℓ, $\ell = 1, \cdots, n$ are independent and identically distributed random variables and X_1 is bounded almost surely. Suppose ϵ_ℓ are independent and identically distributed Gaussian noise $N(0, \sigma^2)$ and $f(X)$ is a bounded linear functional. Then $\widehat{S_{\alpha,\rho,n}}$ converges to $S_{\alpha,\rho}$ in probability with convergence rate in (10).*

5 Approximation of Linear Functionals based on Discrete Observations

In practice, we do not know X completely over the domain $\mathcal{D}$. Instead, we have observations of X over some designed points $s_k, k = 1, .., N$ over $\mathcal{D}$. Let S_X be the discrete least square fit of X assuming that $s_k, k = 1, .., N$ are evenly distributed over $\triangle$ of $\mathcal{D}$ with respect to $S^r_d(\triangle)$. For simplicity, we discuss here the case with no penalty. We consider α_S that solves the following minimization problem:

$$\alpha_S = \arg\min_{\beta \in H} E\left[(f(X) + \epsilon - \langle \beta, S_X \rangle)^2\right]. \tag{16}$$

Also we look for an approximation $S_{\alpha_S} \in S^r_d(\triangle)$ of α_S such that

$$S_{\alpha_S} = \arg\min_{\beta \in S^r_d(\triangle)} E\left[(f(X) + \epsilon - \langle \beta, S_X \rangle)^2\right]. \tag{17}$$

We first analyze how α_S approximates α. It is easy to see that

$$F(\beta) = E\left[(f(X) + \epsilon - \langle \beta, X)^2\right]$$

is a strictly convex function and so is $F_S(\beta) = E\left[(f(X) + \epsilon - \langle \beta, S_X \rangle)^2\right]$. Note that S_X approximates X very well as in Theorem 2.1 as $|\triangle| \to 0$. Thus, $F_S(\beta)$ approximates $F(\beta)$ for each β. Since the strictly convex function has a unique minimizer and both $F(\beta)$ and $F_S(\beta)$ are continuous, α_S approximates α. Indeed, if $\alpha_S \to \beta \neq \alpha$, then $F(\alpha) < F(\beta) = F_S(\beta) + \eta_1 = F_S(\alpha_S) + \eta_1 + \eta_2 \leq F_S(\alpha) + \eta_1 + \eta_2 = F(\alpha_S) + \eta_1 + \eta_2 + \eta_3$ for arbitrary small $\eta_1 + \eta_2 + \eta_3$. Thus we would get the contradiction $F(\alpha) < F(\alpha)$.

We now begin to analyze how S_{α_S} approximates α_S in terms of the size $|\triangle|$ of triangulation. Recall that $\{\phi_1, \cdots, \phi_m\}$ forms a basis for $S^r_d(\triangle)$. We write $S_{\alpha_S} = \sum_{j=1}^m c_{S,j}\phi_j$. Then its coefficient vector $\mathbf{c}_S = (c_{S,1}, \cdots, c_{S,m})^T$ satisfies $A_S\mathbf{c}_S = \mathbf{b}_S$ with A_S being a matrix of size $m \times m$ with entries $E(\langle\phi_i, S_X\rangle\langle\phi_j, S_X\rangle)$ for $i, j = 1, \cdots, m$ and $\mathbf{b}_S$ being a vector of length m with entries $E((f(X) + \epsilon)\langle\phi_j, S_X\rangle)$ for $j = 1, \cdots, m$. We can show that A_S converges to A as $|\triangle| \to 0$ because $E(\langle\phi_i, S_X\rangle\langle\phi_j, S_X\rangle) \to E(\langle\phi_i, X\rangle\langle\phi_j, X\rangle)$ as $S_X \to X$ by Theorem 2.1. That is, we have $\|S_X - X\|_{\infty,\mathcal{D}} \le C|\triangle|^\nu |X|_{\nu,\infty,\mathcal{D}}$ for $X \in W_2^\nu(\mathcal{D})$ with $\nu \ge r > 0$.

To see that S_{α_S} is a good approximation of α_S, we let $\{\phi_j, j = m + 1, m + 2, \cdots\}$ be a basis of the orthogonal complement space of $S^r_d(\triangle)$ in H as before. Then we can write $\alpha_S = \sum_{j=1}^\infty c_{S,j}\phi_j$. Note that the minimization in (16) yields $E(\langle\alpha, S_X\rangle\langle\phi_j, S_X\rangle) = E((f(X)+\epsilon)\langle\phi_j, S_X\rangle)$ for all $j = 1, 2, \cdots$ while the minimization in (17) gives

$$E(\langle S_{\alpha_S}, S_X\rangle\langle\phi_j, S_X\rangle) = E((f(X) + \epsilon)\langle\phi_j, S_X\rangle)$$

for all $j = 1, 2, \cdots, m$. It follows that

$$E(\langle\alpha_S - S_{\alpha_S}, S_X\rangle\langle\phi_j, S_X\rangle) = 0 \tag{18}$$

for all $j = 1, 2, \cdots, m$. Let Q_α be the quasi-interpolatory spline in $S^r_d(\triangle)$ which achieves the optimal order of approximation of α_S from $S^r_d(\triangle)$ as in Preliminary section. Then (18) implies that

$$\begin{aligned} E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) &= E(\langle S_\alpha - S_{\alpha_S}, S_X\rangle\langle\alpha_S - Q_{\alpha_S}, S_X\rangle) \\ &\le (E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2))^{1/2} E((\langle\alpha_S - Q_{\alpha_S}, S_X\rangle)^2)^{1/2}. \end{aligned}$$

It yields $E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) \le E((\langle\alpha_S - Q_{\alpha_S}, S_X\rangle)^2) \le \|\alpha_S - Q_{\alpha_S}\|_H^2 E(\|S_X\|^2)$. The convergence of S_X to X implies that $E(\|S_X\|^2)$ is bounded by a constant dependent on $E(\|X\|^2)$. The approximation of the quasi-interpolant Q_{α_S} of α_S (Theorem 2.1) gives:

Theorem 5.1 *Suppose that $E(\|X\|^2) < \infty$ and suppose $\alpha \in C^r(\mathcal{D})$ for $r \ge 0$. Then the solution S_{α_S} from the minimization problem (17) approximates α_S in the following sense: $E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) \le C|\triangle|^{2r}$ for a constant C dependent on $E(\|X\|^2)$, where $|\triangle|$ is the maximal length of the edges of $\triangle$.*

Next we consider the empirical estimate of S_α based on discrete observations of random surfaces $X_i, i = 1, \cdots, n$. The empirical estimate $\widetilde{S_{\alpha,n}} \in S^r_d(\triangle)$ is the solution of

$$\widetilde{S_{\alpha,n}} = \arg \min_{\beta\in S^r_d(\triangle)} \frac{1}{n}\sum_{i=1}^n (f(X_i)+\epsilon_i - \langle \beta, S_{X_i}\rangle)^2.$$

In fact the solution of the above minimization is given by $\widetilde{S_{\alpha,n}} = \sum_{i=1}^m \widetilde{c_{n,i}}\phi_i$ with coefficient vector $\widetilde{\mathbf{c}_n} = (\widetilde{c_{n,i}}, i=1,\cdots,m)$ satisfying $\widetilde{A_n}\widetilde{\mathbf{c}_n} = \widetilde{b_n}$, and

$$\widetilde{A_n} = \left[\frac{1}{n}\sum_{\ell=1}^n \langle \phi_i, S_{X_\ell}\rangle\langle \phi_j, S_{X_\ell}\rangle\right]_{i,j=1,\cdots,m},$$

where S_{X_ℓ} is the discrete least squares fit of X_ℓ and

$$\widetilde{b_n} = \left[\frac{1}{n}\sum_{\ell=1}^n f(X_\ell)\langle \phi_j, S_{X_\ell}\rangle + \frac{1}{n}\sum_{\ell=1}^n \langle \phi_j, \epsilon_\ell S_{X_\ell}\rangle\right]_{j=1,\cdots,m}.$$

Recall the definition of $\widehat{A_n}$ in Section 3. We have

$$\widetilde{A_n} - \widehat{A_n} = \left[\frac{1}{n}\sum_{\ell=1}^n \langle \phi_i, S_{X_\ell}\rangle\langle \phi_j, S_{X_\ell}\rangle - \frac{1}{n}\sum_{\ell=1}^n \langle \phi_i, X_\ell\rangle\langle \phi_j, X_\ell\rangle\right]_{i,j=1,\cdots,m}.$$

As S_{X_ℓ} converges X_ℓ as $|\triangle|\to 0$, i.e., $S_{X_\ell} - X_\ell = O(|\triangle|^\nu)$, we can show that $\|\widetilde{A_n} - \widehat{A_n}\|_\infty = O(|\triangle|^{\nu-2})$ and hence, $\|\widetilde{A_n} - \widehat{A_n}\|_\infty \to 0$ if $\nu \geq 3$. Likewise, $\widetilde{b_n} - \widehat{b_n}$ converges to 0. We leave the details to the interested reader. Then Lemma 3.2 implies that $\widetilde{S_{\alpha,n}}$ converges to $\widehat{S}_{\alpha,n}$ as $|\triangle|\to 0$ under certain assumptions on $X_\ell, \ell = 1,\cdots,n$ with $n > m$ and $\nu > 4$. Indeed, let us assume that the surfaces $X_\ell, \ell=1,\cdots,n$ are orthonormal and span a space which contains $S^r_d(\triangle)$ or form a tight frame of a space which contains $S^r_d(\triangle)$. Then we can show that the condition numbers $\kappa(\widehat{A_n})$ are bounded by n. Note that the condition number of $\kappa(\widehat{A_n})$ are dependent on the largest and smallest eigenvalues of the matrix $\widehat{A_n}$. They are equivalent to

$$\max_{S\in S^r_d(\triangle)} \frac{1}{n\|S\|_2^2}\sum_{\ell=1}^n |\langle S, X_\ell\rangle|^2 \leq \frac{1}{n}\sum_{\ell=1}^n \|X_\ell\|_2^2 = 1$$

and

$$\min_{S\in S^r_d(\triangle)} \frac{1}{n\|S\|_2^2}\sum_{\ell=1}^n |\langle S, X_\ell\rangle|^2 = \frac{1}{n}$$

Let us further assume that $n = Cm$ for some fixed constant $C > 1$. Next we note that the dimension of $S^r_d(\triangle)$ is strictly less than $\frac{d+2}{2}N$ with N being the number of triangles in $\triangle$ while N can be estimated as follows. Let $A_{\mathcal{D}}$ be the area of the underlying domain $\mathcal{D}$ and assume that the triangulation $\triangle$ is quasi-uniform (cf. [16]). Then $N \le C_1 A_{\mathcal{D}}/|\triangle|^2$ for a positive constant C_1. Thus, the condition number $\kappa(\widehat{A_n}) \le Cm \le CC_1 A_{\mathcal{D}}|\triangle|^{-2}$ when $\nu > 4$. Therefore, Lemma 3.2 implies that the coefficients of $\widetilde{S_{\alpha,n}}$ converges to that of $\widehat{S}_{\alpha,n}$ as $|\triangle| \to 0$. With Lemma 3.1, we conclude that $\widetilde{S_{\alpha,n}}$ converges to $\widehat{S}_{\alpha,n}$.

A similar analysis can be carried out for the approximation with a penalized term as in Section 4. The details are omitted here. Instead, we shall present the convergence based on our numerical experimental results in the next section.

6 Numerical Simulation and Experiments

6.1 Simulations

In this subsection, we present a simulation example on a complicated domain, delimited by the United States frontiers, which has been scaled into $[0,1] \times [0,1]$, see Figure 1. With bivariate spline functions, we can easily carry out all the experiments.

We illustrate the consistency of our estimators using the linear functional: $Y = \langle g, X \rangle$ with known function $g(x,y) = \sin(2\pi(x^2+y^2))$ over the (scaled) US domain. The purpose of the simulation is to estimate g from the value Y based on random surfaces X. The bivariate spline space we employed is $S^1_5(\triangle)$, where $\triangle$ consists of 174 triangles (Fig. 1).

We choose a sample size $n = 5, 20, 100, 200, 500$ and 1000. For each $i = 1, \cdots, n$, we first randomly choose a vector $\mathbf{c}_i$ of size m which is the dimension of $S^1_5(\triangle)$. This coefficient vector $\mathbf{c}_i$ defines a spline function S_i. We evaluate S_i over the (scaled) locations of 969 EPA stations around the USA and add a small noise with zero mean and standard deviation 0.4 at each location. We compute a least squares fit $\widetilde{S}_i$ of the resulting 969 values by using the spline space $S^1_5(\triangle)$ and compute the inner product of g and $\widetilde{S}_i$. We add a small noise of zero mean and standard deviation 0.0002 to get a noisy value Y_i of the functional. Secondly we build the associated matrix $\widetilde{A_n}$ as in section 5 and the right-hand side vector $\widetilde{b_n}$. Finally we solve it to get the solution

Table 1: Errors for the differences $\widetilde{S_{\alpha,\rho,n}} - S_\alpha$ for the simulation and sample sizes n =5, 20, 100, 200, 500 and 1000 based on 20 Monte Carlo simulations and 174 triangles.

sample size		L^2error	
	min	mean	max
n =5	0.671	2.195	31.821
n =20	0.427	0.564	0.666
n =100	0.080	0.115	0.153
n =200	0.048	0.060	0.081
n =500	0.036	0.040	0.044
n =1000	0.029	0.032	0.035
sample size		L^∞ error	
	min	mean	max
n =5	1.242	1.988	3.086
n =20	1.398	2.221	3.584
n =100	0.336	0.468	0.717
n =200	0.158	0.254	0.534
n =500	0.112	0.136	0.207
n =1000	0.092	0.102	0.123

vector $\mathbf{c}$ and spline approximation $\widetilde{S_{g,n}}$ of g. We then evaluate g and $\widetilde{S_{g,n}}$ at locations which are the 101×101 equally spaced points over $[0,1] \times [0,1]$ that fall into the US domain, to compute their differences and find their maximum as well as L_2 norm. We carry out a Monte Carlo experiment over 20 different random seeds. The numerical results show that we approximate well the linear functional, see Table 1. An example of $S_{g,500}$ is shown in Fig. 2.

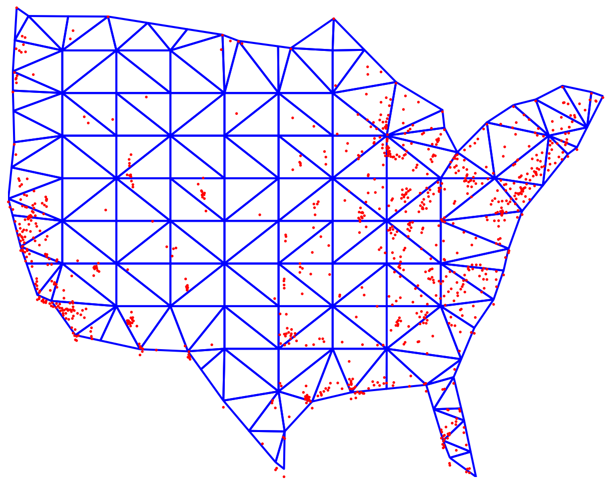

Figure 1: Locations of EPA stations and a Triangulation

6.2 Ozone concentration Forecasting

In this application, we forecast the ground-level ozone concentration at the center of Atlanta using the random surfaces over the entire U.S. domain based on the measurements at various EPA stations from the previous days. Assume that the ozone concentration in Atlanta on one day at a particular time is a linear functional of the ozone concentration distribution over the U.S. continent on the previous day. Also we may assume that the linear functional is continuous. These are reasonable assumptions as the concentration in Atlanta is proportional to the concentration distribution over the entire U.S. continent and a small change in the concentration distribution over the U.S. continent results a small change of the concentration at Atlanta under a normal circumstance. Thus, we build one regression model of the type (1), where $f(X)$ is the ozone concentration value at the center of Atlanta at one hour of one day and X is the ozone concentration distribution function over entire U.S. continent at the same hour but on the previous day, and g is estimated using the penalized least squares approximation with penalty $(= 10^{-6})$ presented in the previous section. Mainly we use a penalized least squares fit S_X of X instead of the discrete least squares fit in the previous section to carry out the empirical estimate $\widetilde{S_{\alpha,n}}$ for S_g.

Let us explain our experiment in detail. To forecast the ozone concen-

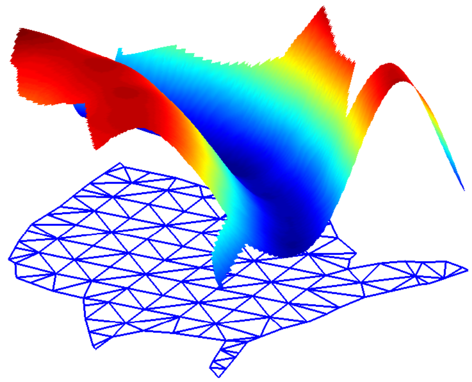

Figure 2: The Surface of Spline Approximation $S_{g,n}$

tration, say on Sept. 12, 2005 at Atlanta, we use the measurements over 14 days before Sept. 11 to build random surfaces X and use the values of the functional Y over 14 days before Sept. 12, so that we can compute an approximation S_g. Once we get S_g, we use it to predict the ozone concentration at the ground-level at Atlanta on Sept. 12 based on $\langle S_g, X\rangle$, where X is the random surfaces based on the measurements on Sept. 11. That is, the prediction $\langle S_g, X\rangle$ is obtained based on a 14 day learning period. Similarly, we can make predictions based on a learning period of a few or more days than 14 days. We show the prediction values on five diferent days, together with the measurements based on 13 to 20 days learning periods. It is easy to see that our spline predictions are very closed to the true measurements. See more experimental results in [8].

This may be compared with the univariate functional autoregressive ozone concentration prediction method [9], but here with no exogenous variables. The idea is to consider a time series of functions which correspond to the ozone concentrations at the location of interest over 24 hours, and then build an autoregressive Hilbertian (ARH) model for this time series. The estimation of the autocorrelation operator in a reduced subspace enables predictions. We selected only 5 functional principal components in the dimension reduction process to keep parsimony in our model, due to sample sizes of 13 to 20. As we see on Figures 3 to 7, the forecasts provided by the 2-D spline

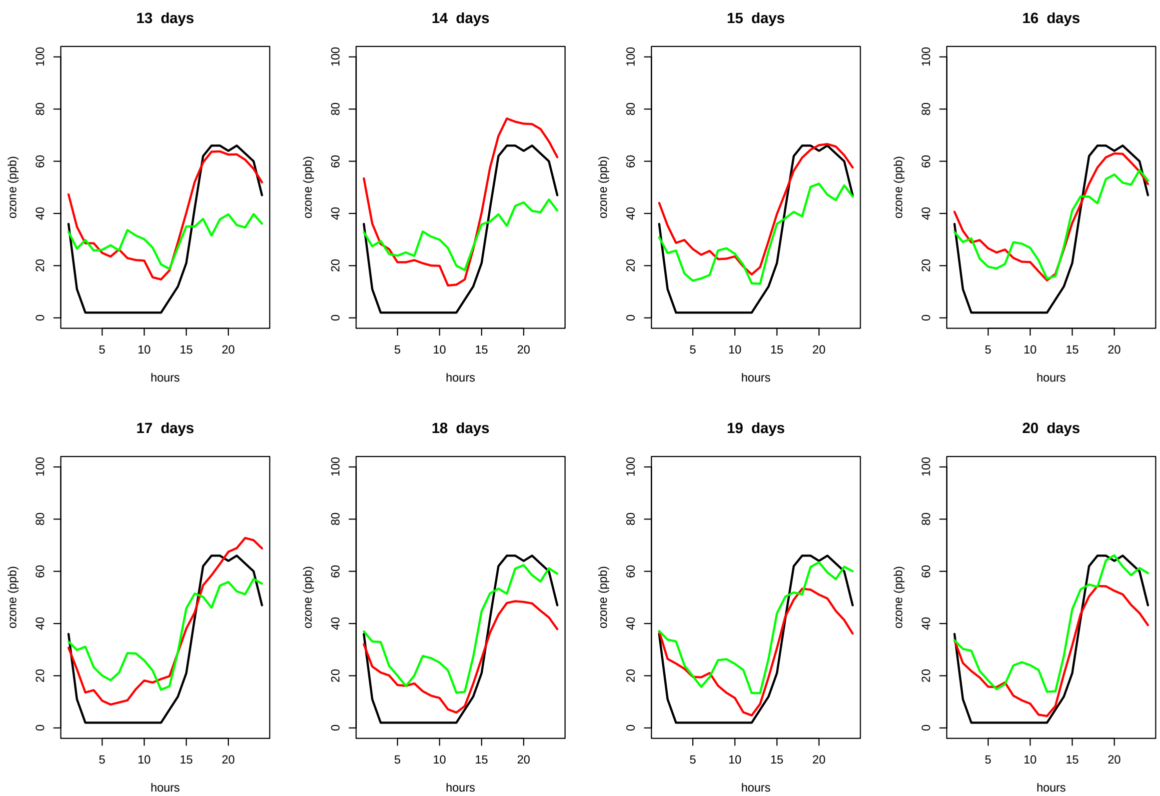

Figure 3: Ozone concentrations in Atlanta on Sept. 9, 2005. Observations (black), forecast 1-D (red), forecast 2-D (green)

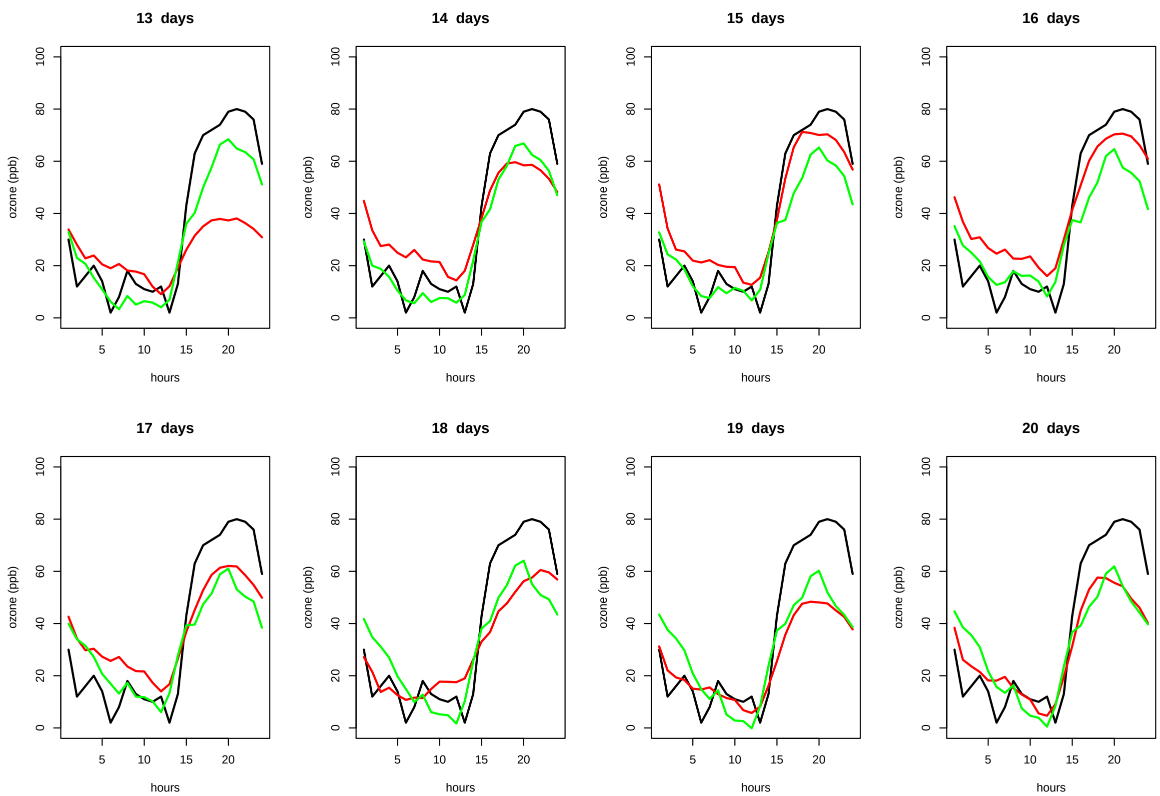

Figure 4: Ozone concentrations in Atlanta on Sept. 10, 2005. Observations (black), forecast 1-D (red), forecast 2-D (green)

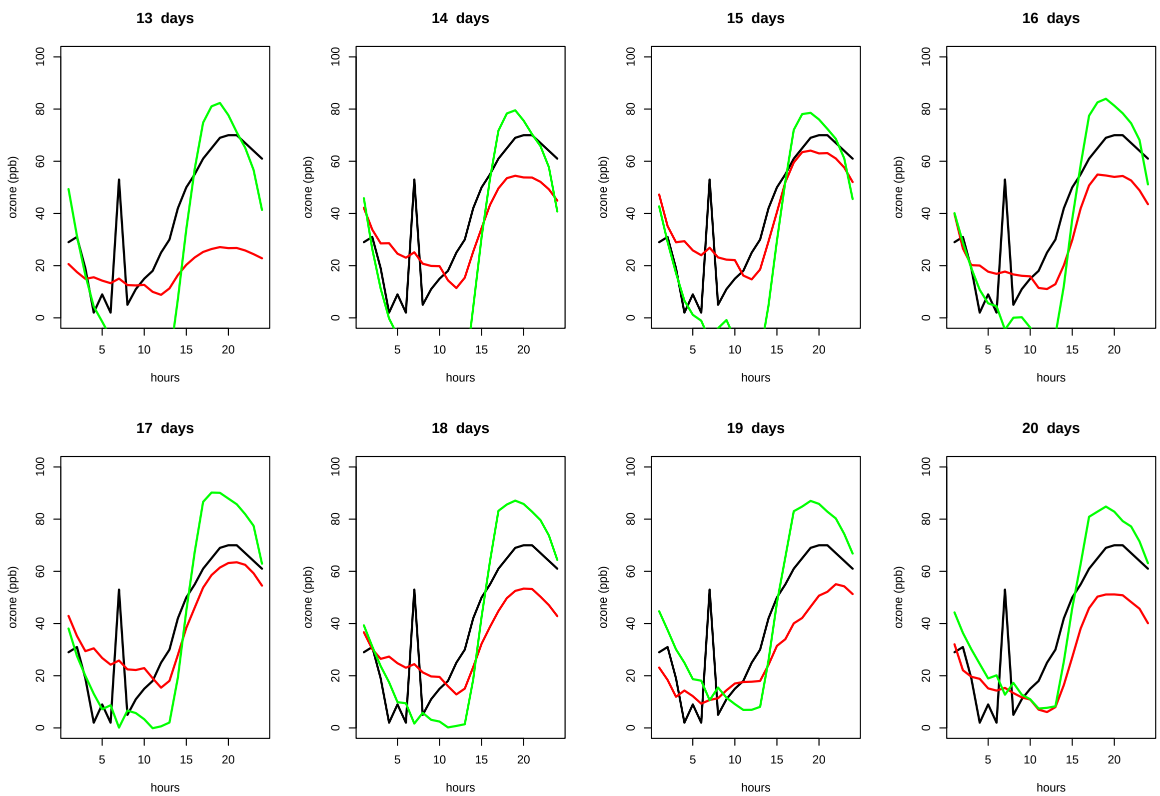

Figure 5: Ozone concentrations in Atlanta on Sept. 11, 2005. Observations (black), forecast 1-D (red), forecast 2-D (green)

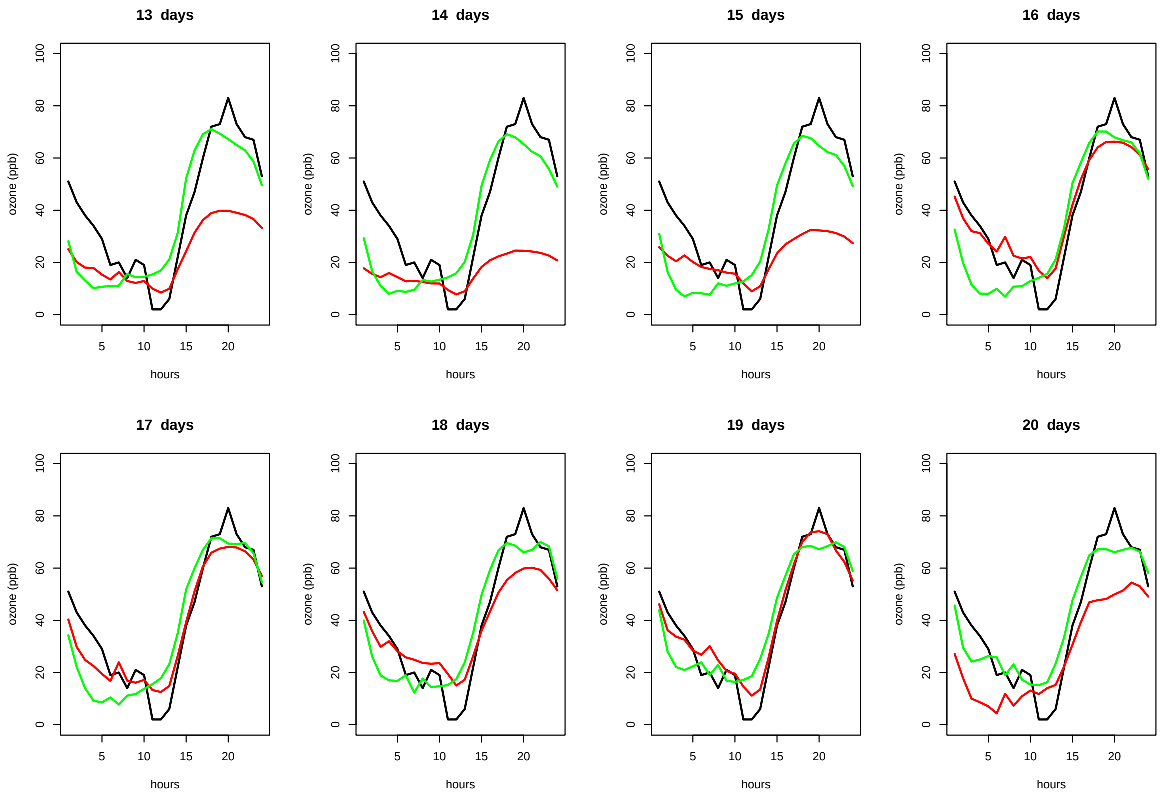

Figure 6: Ozone concentrations in Atlanta on Sept. 12, 2005. Observations (black), forecast 1-D (red), forecast 2-D (green)

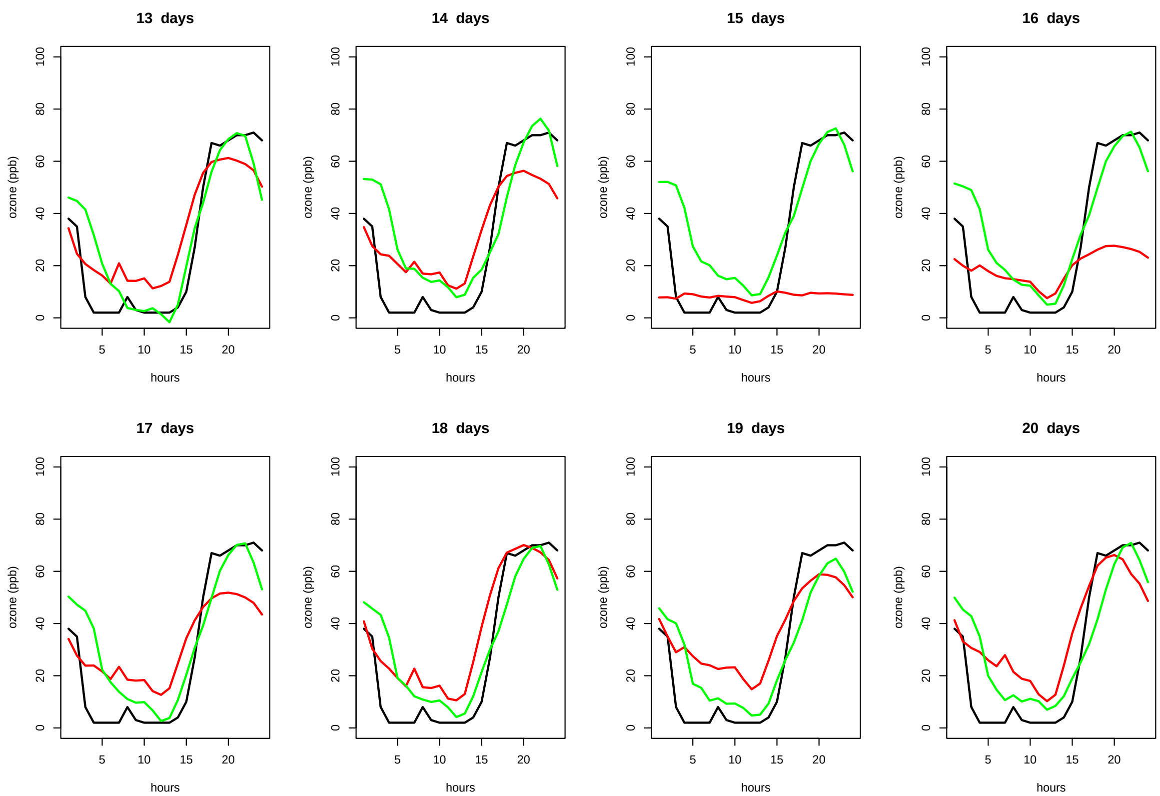

Figure 7: Ozone concentrations in Atlanta on Sept. 13, 2005. Observations (black), forecast 1-D (red), forecast 2-D (green)

strategy outperforms the univariate functional autoregressive method based on the same sizes of samples; only on September 9 the 2-D method did not predict better for small sample sizes, but did better for large sample sizes. This may be explained by the fact that the 2-D approach uses more information to construct its forecasts. We acknowledge that our 2-D approach wrongly assumes that the predictors are independent, whereas the 1-D approach explicitly includes the overall day-to-day dependence. Nevertheless, the comparisons show that our bivariate spline technique almost consistently predicts the ozone concentration values which are closer to the observed values for these 5 days for various learning periods, especially near the peaks. The 1-D method presented in this paper which is considered to be among the best of many 1-D forecasting methods [9] in the following senses: the 1-D method is not consistent for various learning periods and the patterns based on the 1-D method are not as close to the exact measurements as those based on the bivariate spline method in particular in Fig. 4, 6, and 7.

Acknowldgement: The authors would like to thank Ms Bree Ettinger for help in performing the numerical experiments using 2D splines for ozone concentration predictions reported in this paper.

The second author is pleased to acknowledge the support by National Science Foundation under grant #DMS 0713807.

References

[1] Awanou, G. and Lai, M. J. and Wenston, P., The multivariate spline method for numerical solution of partial differential equations and scattered data interpolation, in Wavelets and Splines: Athens 2005, G. Chen and M. J. Lai (eds), Nashboro Press, 2006, 24–74.

[2] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Volume 110 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

[3] Cai, T. T. and Hall, P., Prediction in functional linear regression, Ann. Stat., (2007), to appear.

[4] Cardot, H. and Ferraty, F. and Sarda, P., Functional linear model, Stat. Probab. Lett., (1999) 45, 11–22.

[5] Cardot, H. and Ferraty, F. and Sarda, P., Spline estimators for the functional linear model, Stat. Sin., 2003, 13, 571-591.

[6] Cardot, H. and Crambes, C. and Sarda, P., Spline estimation of conditional quantiles for functional covariates., C. R. Math., (2004),339, 141-144.

[7] Cardot, H. and Sarda, P., Estimation in generalized linear models for functional data via penalized likelihood, J. Multivar. Anal., 92 (2005), 24-41.

[8] Bree Ettinger, Bivariate Splines for Ozone Density Predictions, Dissertation (under preparation), Univ. of Georgia, Aug. 2009.

[9] Damon, J. and S. Guillas (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics 13*, 759–774.

[10] Ferraty, F. and Vieu, P., *Nonparametric Functional Data Analysis: Theory and Practice*, Springer-Verlag, London, 2006.

[11] Golub, Gene H. and Van Loan, Charles F., *Matrix computations*, Johns Hopkins University Press, Baltimore, MD., 1989.

[12] Hall, P. and Horowitz, J. L., Methodology and convergence rates for functional linear regression, Ann. Stat., 2007, to appear.

[13] Hall, P. and Muller, H. G. and Wang, J. L., Properties of principal component methods for functional and longitudinal data analysis, Ann. Stat., (2006), 34, 1493-1517.

[14] Lai, M.-J., Multivariate splines for data fitting and approximation, in *Approximation Theory XII: San Antonio 2007*, M. Neamtu and L. L. Schumaker (eds.), Nashboro Press (Brentwood), 2007, 210–228.

[15] Lai, M.-J., Schumaker, L. L., On the approximation power of bivariate splines; Advances Comput. Math.; 9 (1998) 251–279.

[16] Lai, M.-J., Schumaker, L. L., *Spline Functions on Triangulations*, Cambridge University Press (Cambridge), 2007.

[17] Ramsay, J. and Silverman, B.W., *Functional Data Analysis*, Springer-Verlag, 2005.

[18] Ramsay, T., Spline smoothing over difficult regions, J. R. Stat. Soc. Ser. B Stat. Methodol., 64, (2002), 307–319.

[19] Wood, S. N., Thin plate regression splines, J. R. Stat. Soc. Ser. B Stat. Methodol., 65, (2003), 95–114.

[20] Yao, F. and Lee, T. C. M., Penalized spline models for functional principal component analysis, J. R. Stat. Soc. Ser. B-Stat. Methodol., (2006), 68, 3-25.